

Durham Research Online

Deposited in DRO:

24 February 2011

Version of attached file:

Accepted Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Tymms, P. (2011) 'Evidence? The impact of large-scale reform in England.', *Zeitschrift für Erziehungswissenschaft.*, 13 (S1). pp. 105-115.

Further information on publisher's website:

<http://dx.doi.org/10.1007/s11618-010-0155-2>

Publisher's copyright statement:

The original publication is available at www.springerlink.com

Additional information:

Special edition: Transforming Education = Umbau des Bildungswesens.

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

Tymms, P. (2011). "The Impact of Large-Scale Reform in England." Zeitschrift für Erziehungswissenschaft 13 105-116

Evidence? The Impact of Large-Scale Reform in England

Peter Tymms

Abstract

England's school based education system has been massively reformed over the last two decades. The reforms have included the introduction of a National Curriculum, national testing, a heavy inspection regime and hundreds of lesser reforms.

What has the impact of those reforms been? This paper draws on extensive research and concludes that one of the main aims of the reforms, to impact on basic skills in primary schools, has largely not been achieved although there has been some rise in maths test scores. On the other hand the numbers of students gaining certificates at the age of 16 has more than doubled and more students are studying for pre-university courses and staying in education to study for a degree. The rise in the numbers gaining qualifications at 16 seems to have been achieved whilst maintaining the standards of the exams at least up to 2003 although the same cannot be said of the pre-university qualifications.

The paper concludes that there is a need to take much more notice of the evidence base for reform, to become a learning society and to accept that we need to change policies in the light of evidence.

Key words: Reform, England, Standards, Evidence, Tests

Zusammenfassung:

Evidenzen? Die Auswirkungen groÙ angelegter Reform in England

Das englische Schulsystem ist in den vergangenen zwei Jahrzehnten mächtig reformiert worden. Die Reformen beinhalteten die Einführung eines nationalen Curriculums, nationale Testverfahren, ein starkes Inspektionssystem und hunderte kleinerer Reformen. Was waren die Auswirkungen dieses Reformen? Mein Beitrag stützt sich auf eingehende Untersuchungen und kommt zu dem Schluss, dass eines der Hauptziele der Reformen - die Erhöhung der Leistungen in den Grundfähigkeiten im Primarschulsystem - weitgehend unerreicht blieb, auch wenn es einige Steigerungen in den Mathematikleistungen gab. Auf der anderen Seite hat sich die Zahl der Schülerinnen und Schüler, die mit 16 Jahren ein Abschlusszertifikat erreichen, mehr als verdoppelt, und die Zahl derjenigen, die universitätsvorbereitende Kurse wahrnehmen und einen Hochschulabschluss erreichen möchten, hat sich ebenfalls erhöht. Dies scheint (zumindest bis 2003) erreicht worden zu sein, ohne dass Standards gelockert wurden; allerdings gilt das nicht für die Hochschulzulassungsqualifikationen.

Mein Beitrag kommt zu dem Schluss, dass den Evidenzen, die Reformen zugrunde gelegt werden, mehr Beachtung geschenkt werden muss, um eine lernende Gesellschaft zu werden - und dass es im Lichte der Evidenzen zu einer Änderungen der Politik kommen muss.

Schlüsselwörter: Reform, England, Standards, Evidenz, Tests

Introduction

This paper outlines the major reforms that have defined the major educational changes within England since the end of the 1980s until the present time. It summarises their impact on attainment in state schools during that time, and some of the unintended consequences of the reforms. It also notes the cost of the policies that have been put into place and the recent changes that have been made to the initial reforms in the light of the way things have evolved. The paper finishes up by drawing some lessons from what has happened in England, not just in terms of the reforms themselves and their impact?, but also in terms of the methodology that is used to evaluate the those reforms, and also the basis of the reforms themselves.

Jim Callaghan, the British Prime Minister 1976-1979, became alarmed when he heard that his grandchild was apparently not being taught to read in primary school. He heard this from his daughter and it generated sufficient concern in him to alert the nation to what were perceived to be problems in the educational system in what is now known as the Ruskin College Speech (Callaghan, 1976). As intended this precipitated an extensive debate in England and as a result the Education Reform Act was passed in 1988 which came into force in September 1989.. This put a National Curriculum in place for the first time in England and associated with that was a National Testing Programme for 7, 11, and 14 year olds and the end of what were known as Key Stages, as well as a very high-stakes inspection system known as Ofsted (the Office for Standards in Education). It not only required regular inspections but the inspections themselves could not be challenged in law and their reports would be available through public dissemination of the findings. The test results would also be publicly available and in due course appeared in the newspapers in the form of league tables. These were designed as high-stakes reforms and were meant to hold schools to account for what they were doing after defining what they should be doing.

Tests and exams

The statutory National Testing took some time to become established according to the Secretary of State for Education at the time they were officially bedded down by 1995 and started to come on-stream providing national test data year on year. They added to existing optional assessments for children when they left school, the General Certificate of Secondary Education (GCSE), which had itself taken over in 1986 from two previous parallel testing systems known as GCE O-Levels (General Certificate of Education – Ordinary Level), and the CSE (Certificate of Secondary Education). They catered for the more and less able pupils respectively. At the time of the Education Reform Act (1988) the GCSE was taken by about a quarter of pupils in England. Two years after the GCSE students could take the pre-university assessment known as the A-Levels, the Advanced Level system, which had been in place since 1951. In 1987 an additional intermediate exam known as the AS-Level (Advanced Supplementary) was introduced which was designed to represent a qualification between GCSE and A-Level. There was also a patchwork of assessments of children when they started school using what is known as Baseline Assessment. In England typically children start school in September at the age of 4 although by law they only have to start after their fifth birthday provided they are not being educated at home.

Continued reform

The reforms were initiated by the leader of the Labour Party, Jim Callaghan,, but when the Conservative Party gained power in 1979 they continued the reforms with zeal and it was the Tories who introduced the NC quite quickly after they came to power. Then when Labour came to office in 1997 they also extended the move for reforms using the mantra of *Education, Education, Education*. Indeed in their first few years in office they put in more than 600 initiatives to do with basic skills alone in the primary years (ref). They introduced the National Numeracy Strategy and the National Literacy Strategy in 1998 which were later to become the National Strategy. They brought in target setting, homework clubs, and a whole range of initiatives designed to raise attainment. The focus was on basic skills in primary schools.

The reforms continue and new initiatives and major official reviews continue up to the point of writing this paper.

Standards in primary schools

In Figure 1 the results of the national tests for the end of primary school (age 11 years at the end of Key Stage (KS) 2) are shown. It gives the percentage of children attaining what is known as the “expected”¹ level, Level 4, from 1995 onwards. This shows a dramatic rise between 1995 and 2000 and then a flattening off. At first sight this appears to be major evidence for the positive impact of the reforms, although, as any trained educational researcher knows, it is impossible to ascribe cause with any certainty from such time series data since so many different things were happening simultaneously. But it was certainly used as justification for the reforms that had been put in place. Indeed one high profile government adviser and architect of some of the reforms, Michael Barber claimed that “*Large scale reform is not only possible but can be achieved quickly*”. (Barber 2000, see also Cheek, Fitz-Gibbon and Tymms 2000).

Figure 1 about here

On closer examination the test results seem a little odd. Why is it that they rise dramatically up to 2000 and are then more or less flat thereafter? Why is it that the maths results hug the English results so tightly with a slight discrepancy in 1998 when an oral arithmetic test was introduced? Surely the reforms would be more effective in either English or mathematics. The Centre for Evaluation and Monitoring (CEM, www.cemcentre.org; Tymms and Coe 2003), the largest educational research group in a UK university, had been collecting test results on reading and mathematics over several years using the same test every year, with the same schools and they had not noticed any reading test rise at all (Tymms and Fitz-Gibbon 2001). Informal conversations with Local Authorities that had used the same reading test over the years similarly had not seen a rise but the Authority officials were wary of saying so in public. Clearly there was some suspicion that things were not all that they seemed and it was possible to investigate further (Tymms 2004). Twelve separate studies were identified including one by the government’s own exam body, the Qualifications and Curriculum Authority (QCA), which resulted in the Massey Report (Massey et al 2003). This is distinctive because it used tests from an early date (1995 or 1996) and later date (1999, 2000 or 2001) with randomly equivalent samples of pupils outside

¹ It was initially thought that the ‘average’ pupil would attain L4 but this was later changed to be ‘expected’.

England (Northern Ireland) to check that both tests were equivalent (a similar proportion of pupils reaching the required level would suggest equivalence). Massey et al found that in the cases of both English and maths the tests were not equivalent; the results for English were highly statistically significant, suggesting that the later tests provided an advantage over the earlier tests. Regarding KS2 English Massey et al write:

... these comparisons would lead us to conclude that the standards set in the 1996 and 1999 versions of the KS2 English tests were different (Massey et al 2003: 46)

These results and those from the other independent longitudinal studies using independent instruments were brought together in Tymms (2004). The best estimate is that the attainment levels in English rose very modestly from about 48% to about 57% deserving a Level 4 or above between 1995 and 2000; in mathematics the rise was rather greater (44% to 61%). The English rise amounts to an Effect Size of about 0.25 and the maths rise to about 0.55. After 2000 the English results have risen by a tiny amount ($ES < 0.1$) and the mathematics by a small amount ($ES \sim 0.2$).

We know that whenever tests are introduced on a national basis we see standards rise (see Koretz 2009). So it is of little surprise to educational researchers with an international perspective to see a graph with the general pattern of Figure 1. As is commonly the case, politicians and advisers associated with the reform took a long time to recognise that the apparent rises were illusory and some still do not accept the finding despite independent confirmation that the statutory test results exaggerated the rise (Statistics Commission 2005). What was seen in the English results was the usual rise with the usual causes but this was augmented by a mistake in the standard setting procedure and it is worth just noting what that error was. QCA, despite changing its name, has had the responsibility for the creating of new tests and to ensure that the standards remain constant year on year over the period in question. This involves setting cut-scores corresponding to the various levels against a political background of enormous pressure to see rises. Up to 2000 when the cut-scores were being fixed at crucial yearly meetings it seems that decisions tended to be very slightly biased towards a lower cut-score than was optimal. This amounted to an apparent slight rise in standards. For this to happen one year is quite understandable but the serious error was that each year the intention was only to equate to the results from the last year. No attempt was made to check that the standards were the same as two, three and more years ago. Even though there was an anchor test it was only used to check to the previous year. The consequence was that the small understandable creep in one year was built on every year and this amounted to a serious misrepresentation by 2000. That error was corrected in 2000 and other safeguards were put in place and as a result the standard-setting procedure has been effective since then.

On a broader note more recent work has tried to comment on standards of attainment in primary schools as far back as data will allow (Tymms and Merrell 2007; Tymms, Bolden and Merrell 2008). These studies indicate that reading levels have remained constant in England since a little after the Second World War. Little is known about writing levels over the same period but in mathematics the largest changes have happened since the recent reforms. In science the main debate concerns the primary curriculum so far as it is possible to ascertain standards have remained steady although there is a single worrying paper which suggests that the Piagetian levels of

children starting secondary school has fallen (Shayer Ginsburg and Coe 2008) . The attitudes of pupils to subjects and to school appear to have remained pretty static over the last two decades.

Standards in secondary schools

Turning our attention now to the results at the end of secondary school, Figure 2 shows the proportion of children gaining 5 passes at levels A*-C² at GCSE (Coe and Tymms 2008). A secondary school in England is generally covers the age range 11-16 or 11-18 but in some the students start at 12 or 13. A common feature is that they all include the age at which the GCSE may be taken. It is clear there have been dramatic changes. Around a quarter of students were attaining the 5 passes which are seen as a key attainment level in 1980. Now more than half attain those passes. Independent investigations using CEM data by Robert Coe (Coe et al 2008, Coe and Tymms 2008) using the same developed ability measures shows that that standards used to award GCSE grades has remained fairly constant up until around 2003 and that children of the same ability are generally being awarded the same grades although there is some variation from subject to subject over at least the last twelve years. This suggests that the proportion of students being accredited at the end of compulsory education has more than doubled over approximately the last 20 years whilst examination standards have been maintained. An abrupt rise started in about 1987 just predating the Education Reform Act and the rising curve does not show any noticeable kicks which might correspond to policy or political change. More schools were entering more pupils for the end of statutory education examinations and the proportion accredited for their work is rising.

Figure 2 about here

The rising number of 16 year olds with GCSE passes has had a knock-on effect on A-Level. Figure 3 shows the measure of developed ability of those students starting study in several A-Level subjects. The test was based on the International Test of Developed Abilities (Ottobre 1987) made available to CEM from Education Testing Services. The scores generally start at around 60 points and fall to around 55 then flattening, rising a little from 2001 onwards. Mathematics follows a similar pattern but the scores are a little higher than for the other subjects. It is not surprising to see this kind of pattern because as a higher proportion of the population go on to do A-Levels we would expect that a less and less able group appear in successive cohorts on average. It is gratifying and unexpected to see the slight rise towards the end.

Figure 3 about here

Next the A level grades of students in different subjects with the same developed ability measures when they start are shown in Figure 4. In looking at the chart one needs to recall that the nature of A-levels has changed over the years. The syllabuses have changes, a half-way stage (AS) has been introduced, the nature of the exams has changed and, perhaps of greatest importance, the courses have been modularised. Nevertheless the chart shows a dramatic rise in the grades given to pupils of apparently similar ability. The most noticeable effect is seen in mathematics where in about 1988 it was clear that an E grade would be expected whereas now it is a B/C.

² The highest grade is and A* and this is followed by A , B, C etc

The changing severity of grading of the subjects has resulted in a very high proportion of children getting A grades at A level A2 whereas before it was a small proportion about 10%. This has led to a crisis in the admission of very able students to universities where many present with straight A's and it is impossible for admissions tutors to differentiate at the top level. As a result we have a number of different independent tests making their mark on the English educational scene at the pre-university level.

Figure 4 about here

The cost and impact of reforms

The reforms were enormously expensive and, as noted earlier, very numerous. Four of the larger ones are briefly considered below.

Ofsted

Ofsted (Office for Standards in Education) in its heyday was costing £100 million a year. Its creation involved reforms to the inspection system and it was a harsh regime without a solid evidence base (see for example Fitz-Gibbon 1998). Unusually the impact of this reform has been well studied by Shaw, Newton, and Aitken (2003). They showed that state secondary schools were getting worse results in the years following inspection. Clearly the purpose of inspection was to raise standards and this is a really worrying finding. The Chancellor of the Exchequer, then Gordon Brown, now Prime Minister of England, called the researchers in to his office. Doug Newton was the only one of the authors available at that time. The Chancellor wanted to know if the methodology that they had employed could be used in other areas such as health. Very soon afterwards the enormous amount of money allocated to Ofsted was reduced dramatically. There is now a different form of inspection; one which is less frequent and not so heavy-handed. Despite these changes there is no good independent evidence for the efficacy of Ofsted as it is now constituted. Ofsted itself tends to rely on anecdotal evidence and data on schools with low test results showing improvement. A approach which are hopelessly confounded with the inevitable regression to the mean making it impossible to disentangle claims from counter-claims.

National Numeracy Strategy

The National Numeracy Strategy cost about £500 million and coincided with a modest rise in maths test scores. As noted above it is difficult to disentangle causal relationships from time series data but it is noted that the steady rise in mathematics results do not show and particular jump corresponding to the introduction of the National Numeracy Strategy.

National Literacy Strategy

The National Literacy Strategy cost £500 million and had no detectable impact on reading levels.

National testing

National testing itself was costing £40 per pupil per subject per year, in 2008 (figure released to the author under the Freedom of Information Act). With 600,000 pupils in

an age cohort and 3 subjects to test at the end of Key Stages 2 and 3 the total casts can readily be calculated although that does not include the expense involved within school for preparation and administration. As has been noted above there have been small rises in English standards and slightly higher rises in mathematics in primary school, whereas in secondary schools there have been rises in the proportion of students gaining qualifications. It is hard to gauge the extent to which these changes can be said to be due to testing itself but it is at least possible that increased testing contributed to the small positive change in maths attainment.

Summarising and looking forward

In conclusion it can be said that England has seen major reforms to its educational system worth billions of pounds over the last 20 years. These have resulted in no change in the reading levels of children in English primary schools – a major intention of the reforms - but they are associated with modest rises in mathematics for the same age group. At the end of secondary education it is clear that more students are getting higher qualifications at GCSE and at A-Level, more are also going to university. It is also clear that the bar has been lowered for A-Level but not for GCSE at least until around 2003. Similar grades are being given for less able students.

The lack of any major impact on basic skills in primary schools is sobering and it is equally worrying that so many hundreds of initiatives were introduced without any scientific attempt to evaluate their impact and so we cannot learn from them. It was assumed that their remedies would work. Perhaps the clearest demonstration of the midset associated with the reforms comes from the National Literacy Strategy which was being evaluated during its development with control and experimental schools. But the evaluation was cancelled on a political decision because it was thought to be more important that the program was rolled out nationally. Later a one million pound *post hoc* evaluation contract was given to a Canadian group headed by Michael Fullan, a long standing friend of Michael Barber, the architect of the strategies. The friendly evaluation reports failed to spot the erroneous data being generated by the national tests. It was all too little too late.

There is a clear need to think anew about reform and that must be guided by three key principles. The first is that evidence rather than opinion must hold sway. In England there was little concern about the evidence base of reforms in the 1990s although the “findings” of school effectiveness research, brought together in Teddlie and Reynolds (2000) were much quoted. But it has long been recognised that the associations touted as the key to reform by school effectiveness researchers were just that, associations, and that an evidence base needs to be founded on tight evaluations of initiatives. This realisation started to gain ground and although initially only lip-service was being paid to such a view more effort is being taken to take notice of evidence such as it exists in the UK and beyond.

Secondly we should aim to become a learning society. All new initiatives and policies, even those with an excellent pedigree, need to be tightly and continuously evaluated. It cannot be sufficient to take a well evaluated and successful reform and assume that it will work when it is implemented. Circumstances are always different and continuous monitoring is needed. Further, we should learn from those

evaluations. Not only should policies be based on evidence but policies should be used to create evidence.

Thirdly we should not forget Campbell's paper "Reforms as Experiments" from 1969. No initiative is perfect and there are always alternatives. We should encourage diversity in our educational systems and help politicians to foster diversity and different approaches to reform. It is very unhealthy to insist that a politician should not change his or her mind and perform a U-turn. The future is an unknown land and we need to treat it as such trying out new routes and altering course as need be..

Thanks

Dr. David Bolden gave valuable suggestions on an earlier draft of this and thanks for due to him.

References

Barber, M. (2000, December 13 2000: Vol XX, No 15). *Education Week: American Education's Newspaper of Record*, p. 42 and 45.

Callaghan, J. (1976) The full text of this speech can be viewed at:
<http://education.guardian.co.uk/thegreatdebate/story/0,9860,574645,00.html>
[Accessed: 12th March 2009].

Campbell, D. T. (1969). Reforms as experiments. *American Psychologist*, 24, 409-429.

Cheek, D. W., Fitz-Gibbon, C. T., & Tymms, P. (2000, December 13 2000: Vol XX, No 15). The English Reforms Are Not for Us. *Education Week: American Education's Newspaper of Record*, p. 42 and 45.

Coe, R., & Tymms, P. (2008). Summary of research on changes in educational standards in the UK. In M. Harris (Ed.), *Education Briefing Book 2008: IoD Policy Paper*. London: Institute of Directors.

Coe, R., Serle, J., Barmby, P., Jones, K. and Higgins, S. (2008) *Relative Difficulty of Examinations in Different Subjects* Report for SCORE (Science Community Supporting Education) CEM

Fitz-Gibbon, C. (1998). Ofsted: time to go? *Managing Schools Today*, 7(6), 22-25.

Koretz, D. (2009) Lessons from Test-Based Education Reform in the U.S. ZfE Beiheft Nr. 13, vorläufige Gliederung

Massey, A., Green, S., Dexter, T. and Hammet, L. (2003). *Comparability of national tests over time: KS1, KS2 and KS3 standards between 1996 and 2001. Final Report to QCA of the Comparability Over Time Project*. London: Research

and Evaluation Division of the University of Cambridge Local Examinations Syndicate.

- Ottobre, F. M., & Turnbull, W. W. (1987). The International Test of Developed Abilities: A report on the feasibility study. Report for the International Association for Educational Assessment Princeton New Jersey 08541.
- Statistics Commission (2005). *Measuring Standards in English Primary Schools: Report by the Statistics Commission on an article by Peter Tymms* (No. Statistics Commission Report No 23). London: Statistics Commission.
- Shaw, I., Newton, D. P., Aitkin, M., & Darnell, R. (2003). Do OFSTED Inspections of Secondary Schools Make a Difference to GCSE Results? *British Educational Research Journal*, 29(1), 63-75.
- Shayer, M., Ginsburg, D., & Coe, R. (2006). 30 Years on - an anti-'Flynn effect'? The Piagetian test Volume & Heaviness norms 1975-2003. *British Journal of Educational Psychology*.
- Teddlie, C., & Reynolds, D. (2000). *The International Handbook of School Effectiveness Research*. London: Falmer Press.
- Tymms, P., & Fitz-Gibbon, C. (2001). Standards, Achievement and Educational Performance: A Cause for Celebration? In R. Phillips & J. Furlong (Eds.), *Education, Reform and the State: Twenty-Five Years of Politics, Policy and Practice* (pp. 156-173). London: Routledge Falmer.
- Tymms, P. (2004). Are standards rising in English primary schools?, *British Educational Research Journal* (Vol. 30, pp. 477-494).
- Tymms, P., & Merrell, C. (2007). *Standards and Quality in English Primary Schools Over Time: the national evidence: Primary Review Research Survey 4/1* (No. ISBN 978-1-906478-01-8). Cambridge: University of Cambridge Faculty of Education.
- Tymms, P., & Coe, R. (2003). Celebration of the Success of Distributed Research with Schools: the CEM Centre, Durham. *British Educational Research Journal*, 29(5), 639-653.
- Tymms, P., Bolden, D., & Merrell, C. (2008). Science in English primary schools: trends in attainment, attitudes and approaches. *Perspectives on Education: Primary Science (Wellcome Trust)*(1), 19-42.

Charts and Tables

Figure 1 Percent of pupil gaining a Level 4 or above in the statutory tests at the end of primary school

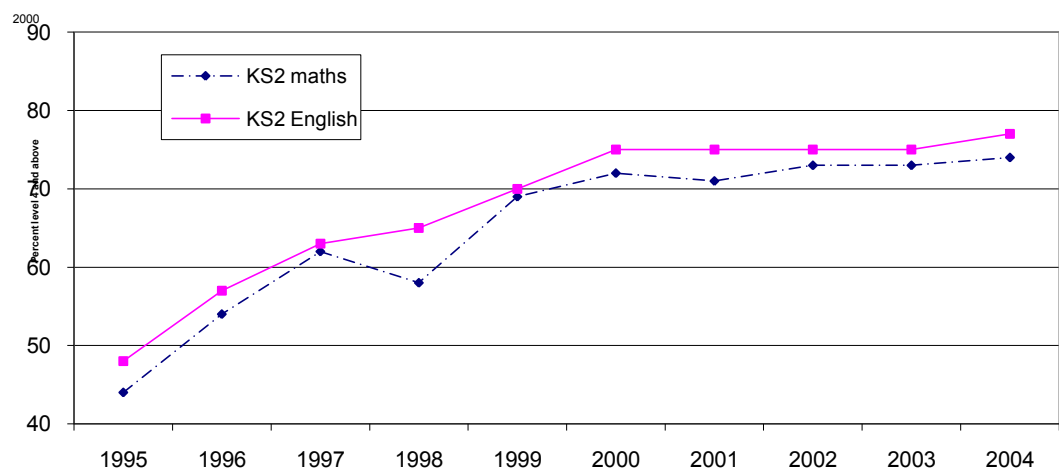


Figure 2 Percent of students gaining five or more good passes at GCSE in England

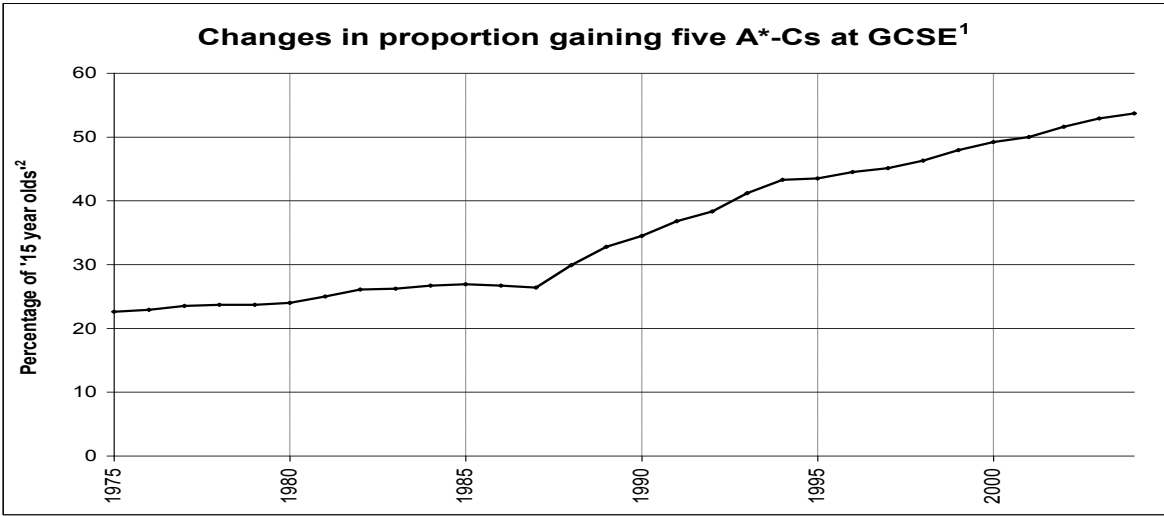


Figure 3 Mean score on a measure of developed ability for students starting A level.

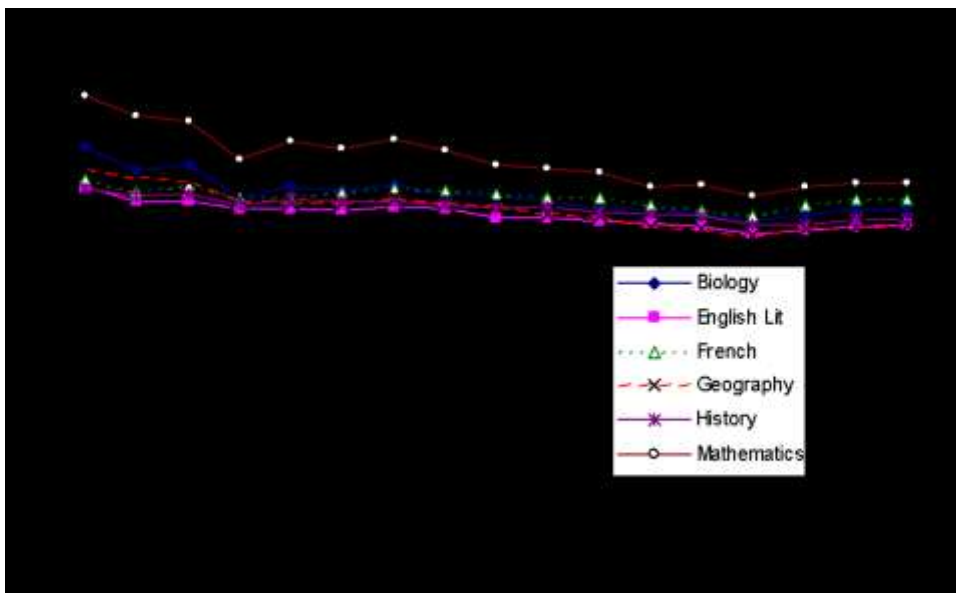


Figure 4 A level grades of students with the same measured developed ability

